*Econometric Theory*, **0**, 2023, 1–24.
doi:10.1017/S0266466623000324

# NONPARAMETRIC TIME-VARYING PANEL DATA MODELS WITH HETEROGENEITY

FEI LIU
*Nankai University*

Since Bai (2009, *Econometrica* 77, 1229–1279), considerable extensions have been made to panel data models with interactive fixed effects (IFEs). However, little work has been conducted to understand the associated iterative algorithm, which, to the best of our knowledge, is the most commonly adopted approach in this line of research. In this paper, we refine the algorithm of panel data models with IFEs using the nuclear-norm penalization method and duple least-squares (DLS) iterations. Meanwhile, we allow the regression coefficients to be individual-specific and evolve over time. Accordingly, asymptotic properties are established to demonstrate the theoretical validity of the proposed approach. Furthermore, we show that the proposed methodology exhibits good finite-sample performance using simulation and real data examples.

## 1. INTRODUCTION

In the panel data model literature, the interactive fixed effects (IFEs) structure has received considerable attention since Pesaran (2006) and Bai (2009). The common correlated effects approach and principal component analysis (PCA) are the main tools used for regression analysis. These approaches are so popular that each has its own separate literature. In this study, we use a PCA-based approach. In terms of numerical implementation, most research has followed Bai (2009) in using an iterative algorithm; however, little work has been done to explore its asymptotic behavior. The only exception known to us is Jiang et al. (2021), who specifically study a homogeneous parametric framework and document that this algorithm suffers from an initial estimation bias. Moon and Weidner (2018) regard the bias as a non-convex minimization problem that can be addressed using a nuclear-norm penalization method. Despite producing consistent estimators, this method exhibits a slow convergence rate.

The author would like to thank the Editor (Peter C.B. Phillips), the Co-Editor (Liangjun Su), and two referees for their constructive comments and suggestions. An earlier version of this paper was a chapter of the author's PhD thesis at Monash University under the supervision of Jiti Gao and Yanrong Yang. The author would like to acknowledge their guidance and helpful comments. Thanks also go to Badi H. Baltagi, Oliver Linton, Bin Peng, and Robin Sickles for their comments on early versions of this paper. This research is financially supported by the National Natural Science Foundation of China under Grant No. 72203114. Address correspondence to Fei Liu, School of Finance, Nankai University, Tianjin, China; e-mail: fei.liu@nankai.edu.cn.

 

In this study, we provide an improved algorithm that can eliminate the initial estimation biases and achieve an optimal convergence rate. Specifically, we employ a semiparametric nuclear-norm penalization method for consistent initial estimation and conduct duple least-squares (DLS) iterations to reduce initial estimation biases. Under mild conditions, we establish asymptotic properties for the initial and iterative DLS estimators. We also consider a generalized model with the IFEs structure by involving regression coefficients that are individual-specific and nonparametric time-varying functions. At this point, we introduce another terminology: profile least squares (PLS), which is an important step in our DLS algorithm. In the literature, the PLS method is commonly adopted for obtaining efficient semi-/non-parametric estimation (Speckman, 1988; Su and Ullah, 2006; Atak, Linton, and Xiao, 2011; Phillips and Wang, 2022; among others). In each iteration of our DLS algorithm, PLS is employed for the simultaneous estimation of time-varying regression coefficients and factor loadings. This results in a regression model for common factors that can be estimated using the OLS method. In summary, the newly proposed DLS algorithm is relatively straightforward in the sense that only two least-squares methods are involved, which is a desirable feature for practical implementation.

Additionally, the present study provides the following outcomes: (1) a residual-based method to test the underlying time-constant parameter assumption versus local alternatives of time-varying functions and (2) an information criterion for factor number selection. Notably, although numerous model specification tests have been proposed in time series and panel data frameworks (e.g., Wooldridge, 1992; Hardle and Mammen, 1993; Li, 1999; Su, Jin, and Zhang, 2015), virtually no studies have tested the time-constant parameter assumption for the IFEs structure. Therefore, we construct a test statistic using restricted estimation residuals of a time-constant IFEs model with heterogeneity and develop asymptotic properties accordingly. The information criterion is established under a set of mild conditions and also works effectively when the heterogeneous coefficient functions reduce to constant parameters. Consequently, it nests Bai and Ng's (2002) method as a special case. In numerical studies, these theoretical findings are examined through extensive simulations.

The remainder of this paper is organized as follows. Section 2 introduces the heterogeneous time-varying panel data model and proposes the DLS estimation method. Section 3 establishes the asymptotic properties of the DLS estimators under regularity conditions. Section 4 develops a constancy test for the time-constant parameter assumption and employs an information criterion for factor number selection. Simulation studies are presented in Section 5. Section 6 applies the newly proposed methodology to an empirical study that investigates the mutual fund performance in the United States. Section 7 concludes the study. The Appendix provides justifications for assumptions and outlines the theoretical development of this study. Additional discussion and proofs are provided in the Supplementary Material.

Before proceeding further, we introduce the following notation: for a matrix $A$, $A^\top$ denotes its transpose; $\|\cdot\|_F$, $\|\cdot\|_*$, and $\|\cdot\|_\infty$ denote Frobenius, nuclear, and spectral matrix norms, respectively; $\|\cdot\|$ denotes $L_2$ vector norm; $\mathrm{tr}(\cdot)$, $\mathrm{rank}(\cdot)$, $\mathrm{vec}(\cdot)$, and $\lambda_{\min}(\cdot)$ denote trace, rank, vectorization, and the smallest eigenvalue of a matrix, respectively. Further, for matrices $A$ and $B$ with the same dimensions, $A \odot B$ denotes their elementwise product matrix; $I_a$ and $0_a$ are $a \times a$ identity and null matrices, respectively; $0_{a\times b}$ is an $a \times b$ matrix of zeros; for an $a \times b$ matrix $A$ with full column rank, let $\mathcal{M}_A = I_a - A(A^\top A)^{-1}A^\top$; $a_n \asymp b_n$ says $a_n$ and $b_n$ have the same order as $n \to \infty$; $\xrightarrow{P}$ denotes convergence in probability, and $\xrightarrow{D}$ denotes convergence in distribution; $K(\cdot)$ and $h$ are the kernel function and bandwidth in the kernel estimation, respectively.

## 2. MODEL AND DLS ALGORITHM

This section presents the model setup and the DLS algorithm. Specifically, we consider the following panel data model:

$$y_{it} = x_{it}^\top \beta_{it} + \lambda_i^{0\top} f_t^0 + \varepsilon_{it}, \quad i = 1,2,\ldots,N,\ t = 1,2,\ldots,T, \tag{2.1}$$

where $x_{it}$ is a $p \times 1$ vector of explanatory variables, $\beta_{it} = \beta_i(\tau_t)$ is a $p \times 1$ vector of unknown nonparametric functions for $\tau_t = t/T$, $\lambda_i^0$ and $f_t^0$ are $r_0 \times 1$ vectors of unobserved factor loadings and common factors, respectively, and $\varepsilon_{it}$ is an error component having correlation along both dimensions. As always, $\{x_{it}\}$ and $\{\lambda_i^0, f_t^0\}$ can be correlated. For the time being, we suppose that the number of factors ($r_0$) is known and consider its estimation in Section 4.2.

### 2.1. Initial Estimation

For notational simplicity, let $B(\tau) = (\beta_1(\tau),\ldots,\beta_N(\tau))^\top$, $\Lambda^0 = \left(\lambda_1^0,\ldots,\lambda_N^0\right)^\top$, $F^0 = \left(f_1^0,\ldots,f_T^0\right)^\top$, $\gamma_{it}^0 = \lambda_i^{0\top} f_t^0$, and $\Gamma^0 = \Lambda^0 F^{0\top}$. The local linear method, as a conventional nonparametric approach (see Chapter 2 of Fan and Gijbels, 1996), can be applied to estimate time-varying functions. Intuitively, for any given $\tau \in (0,1)$, we should consider the following objective function:

$$Q_\tau(A,C,\Gamma) = \frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - x_{it}^\top (a_i + (\tau_t - \tau)c_i) - \gamma_{it}\right)^2 K\left(\frac{t-\tau T}{Th}\right), \tag{2.2}$$

for $A, C \in \mathbb{R}^{N\times p}$ and $\Gamma \in \mathbb{R}^{N\times T}$, where $a_i$ and $c_i$ are the $i$th columns of $A^\top$ and $C^\top$, respectively, and $\gamma_{it}$ is the $(i,t)$th element of $\Gamma$. However, $Q_\tau(A,C,\Gamma)$ is not necessarily convex in $\Gamma$, so minimizing (2.2) numerically may not produce consistent estimators. To overcome this problem, we consider the following local linear objective function with the nuclear-norm regularization:

$$\left(\widehat{B}^{(0)}(\tau), \widehat{B}'^{(0)}(\tau), \widehat{\Gamma}_\tau\right) = \underset{A,\,C\in\mathbb{R}^{N\times p},\,\Gamma\in\mathbb{R}^{N\times T}}{\operatorname{argmin}} \left\{Q_\tau(A,C,\Gamma) + \frac{\phi_{NT}}{\sqrt{NT}} \|\Gamma\|_*\right\}, \tag{2.3}$$

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

where $\phi_{NT}$ denotes a regularization parameter that satisfies $\phi_{NT} > 0$ and $\phi_{NT} \to 0$. Various studies have adopted the same treatment to address similar issues (e.g., Chernozhukov et al., 2018; Moon and Weidner, 2018; Ma, Su, and Zhang, 2022).

Let $\widehat{\beta}_i^{(0)}(\tau)$ be the $i$th column of $\widehat{B}^{(0)\top}(\tau)$ and $R_{it} = y_{it} - x_{it}^\top \widehat{\beta}_i^{(0)}(\tau_t)$. Additionally, we define $R_i = (R_{i1}, \ldots, R_{iT})^\top$ and $R = (R_1, \ldots, R_N)^\top$. Subsequently, we employ the PCA method to construct initial estimators for factors and loadings as follows:

$$\frac{1}{NT}\sum_{i=1}^{N} R_i R_i^\top \widehat{F}^{(0)} = \widehat{F}^{(0)} V_{NT}, \quad \widehat{\Lambda}^{(0)} = T^{-1} R \widehat{F}^{(0)},$$

where $V_{NT}$ is an $r_0 \times r_0$ diagonal matrix containing the first $r_0$ largest eigenvalues of the matrix $(NT)^{-1}\sum_{i=1}^N R_i R_i^\top$ in descending order, and $\widehat{F}^{(0)}$ satisfies $T^{-1}\widehat{F}^{(0)\top}\widehat{F}^{(0)} = I_{r_0}$.

Despite the consistency of $\widehat{B}^{(0)}(\tau)$, $\widehat{F}^{(0)}$, and $\widehat{\Lambda}^{(0)}$, as discussed in Section 3, these estimators suffer from substantial shrinkage biases owing to the inefficiency of the nuclear-norm regularization method. Therefore, we propose the DLS algorithm, which iteratively implements OLS estimation for factors and PLS estimation for individual-specific regression coefficients and factor loadings.

## 2.2. DLS Iteration

Before proceeding further, we introduce some new notation to facilitate the development. We denote $n$ as the number of iterations. Accordingly, let $\widehat{B}^{(n)}(\tau)$, $\widehat{F}^{(n)}$, and $\widehat{\Lambda}^{(n)}$ be the estimators of $B(\tau)$, $F^0$, and $\Lambda^0$, respectively, in the $n$th step. Then, $\widehat{\beta}_i^{(n)}(\tau), \widehat{f}_t^{(n)}$, and $\widehat{\lambda}_i^{(n)}$ are defined analogously. In addition, we define $\widetilde{R}_{it}^{(n)} = y_{it} - x_{it}^\top \widehat{\beta}_i^{(n)}(\tau_t)$, $\widetilde{R}_t^{(n)} = (\widetilde{R}_{1t}^{(n)}, \ldots, \widetilde{R}_{Nt}^{(n)})^\top$, and

$$\begin{aligned}
W(\tau) &= \operatorname{diag}\left(K((\tau_1 - \tau)/h),, \ldots, K((\tau_T - \tau)/h)\right), \\
M_i(\tau) &= \begin{pmatrix} x_{i1} & \cdots & x_{iT} \\ x_{i1}(\tau_1 - \tau)/h & \cdots & x_{iT}(\tau_1 - \tau)/h \end{pmatrix}^\top, \\
s_i(\tau) &= [I_p, 0_p][M_i(\tau)^\top W(\tau) M_i(\tau)]^{-1} M_i(\tau)^\top W(\tau), \\
S_i &= (s_i(\tau_1)^\top x_{i1}, \ldots, s_i(\tau_T)^\top x_{iT})^\top.
\end{aligned}$$

We are now ready to present the iteration procedure.

*Step 1*. Find initial estimators $\widehat{B}^{(0)}(\tau)$, $\widehat{F}^{(0)}$, and $\widehat{\Lambda}^{(0)}$ as in Section 2.1.

*Step 2*. With $\widehat{F}^{(n-1)}$, we employ the PLS method to estimate $\lambda_i^0$ and $\beta_i(\tau)$. Note that for given $\lambda_i^0$ and $\tau$, $\beta_i(\tau)$ can be estimated by

$$\begin{aligned}
&\left(\widehat{\beta}_i^{(n)}(\tau, \lambda_i^0), \widehat{\beta'}_i^{(n)}(\tau, \lambda_i^0)\right) \\
&= \underset{a_i, c_i}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_{it} - \lambda_i^{0\top}\widehat{f}_t^{(n-1)} - x_{it}^\top (a_i + (\tau_t - \tau) c_i)\right)^2 K\left(\frac{t - \tau T}{Th}\right).
\end{aligned}$$

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

Simple algebra yields

$$\widehat{\beta}_i^{(n)}(\tau,\lambda_i^0) = [I_p, 0_p]\left[M_i(\tau)^\top W(\tau) M_i(\tau)\right]^{-1} M_i(\tau)^\top W(\tau)\left[y_i - \widehat{F}^{(n-1)}\lambda_i^0\right], \qquad \textbf{(2.4)}$$

where $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})^\top$. Using $\widehat{\beta}_i^{(n)}(\tau,\lambda_i^0)$, we obtain the PLS estimator of factor loadings:

$$\widehat{\lambda}_i^{(n)} = \underset{\lambda_i}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_{it} - x_{it}^\top \widehat{\beta}_i^{(n)}(\tau,\lambda_i) - \lambda_i^\top \widehat{f}_t^{(n-1)}\right)^2,$$

which yields

$$\widehat{\lambda}_i^{(n)} = \left[\widehat{F}^{(n-1)\top}(I - S_i)^\top (I - S_i)\widehat{F}^{(n-1)}\right]^{-1} \widehat{F}^{(n-1)\top}(I - S_i)^\top (I - S_i) y_i. \qquad \textbf{(2.5)}$$

Plugging $\widehat{\lambda}_i^{(n)}$ into (2.4), we finally obtain the PLS estimator of $\beta_i(\tau)$:

$$\widehat{\beta}_i^{(n)}(\tau) = [I_p, 0_p]\left[M_i(\tau)^\top W(\tau) M_i(\tau)\right]^{-1} M_i(\tau)^\top W(\tau)\left[y_i - \widehat{F}^{(n-1)}\widehat{\lambda}_i^{(n)}\right].$$

*Step 3.* Subsequently, we estimate $f_t^0$ by

$$\widehat{f}_t^{(n)} = \left(\widehat{\Lambda}^{(n)\top}\widehat{\Lambda}^{(n)}\right)^{-1} \widehat{\Lambda}^{(n)\top}\widetilde{R}_t^{(n)}.$$

*Step 4.* Repeat Steps 2 and 3 until a certain convergence criterion is satisfied.

The following section investigates the above algorithm's asymptotic properties under a set of mild conditions.

## 3. ASYMPTOTIC RESULTS

The asymptotic results in this section are organized as follows. Theorems 3.1 and 3.2 establish the consistency of the initial and iterative DLS estimators, respectively. Asymptotic distributions are given in Theorem 3.3. Theorem 3.4 studies the asymptotic properties of the mean group (MG) estimator.

We impose a structure for explanatory variables of the form: $x_{it} = g_{it} + v_{it}$ to capture the trending features, where $g_{it} = g_i(\tau_t)$ is a $p \times 1$ vector of unknown trend functions, and $v_{it}$ denotes a stationary error term that allows interaction with common factors. Let $v_t = (v_{1t}, \ldots, v_{Nt})^\top$, $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\top$, $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_T)$, and $\mathcal{D} = \{\lambda_1^0, \lambda_2^0, \ldots\}$. Additionally, define $\widetilde{M}_{i,\tau} = W^{1/2}(\tau) M_i(\tau)$ and $\mathcal{M}_{x,\tau} = \operatorname{diag}(\mathcal{M}_{\widetilde{M}_{1,\tau}}, \ldots, \mathcal{M}_{\widetilde{M}_{N,\tau}})$.

**Assumption 1.**

(i) $\{v_t, f_t^0, \varepsilon_t\}$ are strictly stationary and $\alpha$-mixing across $t$ conditional on $\mathcal{D}$. Let $\alpha_{ij}^{\mathcal{D}}(|t-s|)$ and $\alpha_0^{\mathcal{D}}(|t-s|)$ represent the conditionally $\alpha$-mixing coefficients between $\{v_{it}, \varepsilon_{it}\}$ and $\{v_{js}, \varepsilon_{js}\}$, and between $\{f_t^0\}$ and $\{f_s^0\}$, respectively. Assume that $\alpha_{ij}^{\mathcal{D}}(t) \le \alpha_{ij}(t)$ almost surely and $\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=0}^{T}\left(\alpha_{ij}(t)\right)^{\delta/(4+\delta)} = O(N)$, where $\delta > 0$ is selected such that $\mathbb{E}(\|\omega_{it}\|^{4+\delta}) < \infty$ for $\omega_{it} \in \{v_{it}, \lambda_i^0, f_t^0, \varepsilon_{it}\}$.

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

Let $\alpha^{\mathcal{D}}(t) = \max\{\max_{1\le i,j\le N}\alpha_{ij}^{\mathcal{D}}(t), \alpha_0^{\mathcal{D}}(t)\}$. Assume that $\alpha^{\mathcal{D}}(t) \le \alpha(t)$ almost surely and $\alpha(t) = O(t^{-\theta})$, where $\theta > (4+\delta)/\delta$.

(ii) $\{\varepsilon_{it}\}$ are mean-zero and independent of $\{v_{js}, f_s^0\}$ conditional on $\mathcal{D}$, for any $(i,j,t,s)$. Moreover, $\|\mathcal{E}\|_\infty = O_P(\max\{\sqrt{N},\sqrt{T}\})$ and $\max_{1\le i\le N, 1\le t\le T}\|v_{it}\| = O_P(\log(NT))$.

(iii) Unknown functions $\{\beta_i(\tau), g_i(\tau)\}$ are uniformly bounded and have continuous derivatives of up to the second order for $\tau \in [0,1]$.

(iv) The kernel function $K(\cdot)$ is Lipschitz continuous with a compact support on $[-1,1]$.

(v) The bandwidth $h$ satisfies that $\lim_{T\to\infty} Th^5 < \infty$, $\log(NT)^2 h \to 0$, $\frac{N}{T^2h^2} \to 0$, $\frac{T}{N^2h^2} \to 0$, and $\min\{N,T\}h^2 \to \infty$, as $N,T\to\infty$. The regularization parameter $\phi_{NT}$ satisfies $\min\{N,T\}h^2\phi_{NT}^2 \to \infty$ and $\max\{N,T\}h^2\phi_{NT}^4 \to 0$, as $N,T\to\infty$.

(vi) Let $\mathcal{C}(c) = \{\Gamma \in \mathbb{R}^{N\times T} : \left\|\mathcal{M}_{\Lambda^0}\Gamma\mathcal{M}_{F^0}\right\|_* \le c\left\|\Gamma - \mathcal{M}_{\Lambda^0}\Gamma\mathcal{M}_{F^0}\right\|_*$, almost surely$\}$, for some constant $c > 0$. A constant number $\kappa_c > 0$ exists such that $\mathrm{vec}(\Gamma^\top)^\top \mathcal{M}_{x,\tau}\, vec(\Gamma^\top) \ge \kappa_c vec(\Gamma)^\top vec(\Gamma)$, for any $\Gamma \in \mathcal{C}(c)$ and $\tau \in [0,1]$.

Assumption 1 comprises regularity conditions such as the stationarity, strict exogeneity, weak cross-sectional dependence and serial correlation of errors, the smoothness of time-varying functions, and the restricted strong convexity for the nuclear-norm penalization. As these assumptions are conventional in the literature, we provide their justifications in Appendix A.1.

Let $\Sigma_f^{\mathcal{D}} = \mathbb{E}_{\mathcal{D}}(f_t^0 f_t^{0\top})$, where $\mathbb{E}_{\mathcal{D}}(\cdot)$ denotes the expectation conditional on $\mathcal{D}$. Additionally, let $\Sigma_{x,i}^{\mathcal{D}}(\tau_t) = \mathbb{E}_{\mathcal{D}}(x_{it}x_{it}^\top)$, $\Sigma_{xf,i}^{\mathcal{D}}(\tau_t) = \mathbb{E}_{\mathcal{D}}(x_{it}f_t^{0\top})$, and $\Omega_{f,i}^{\mathcal{D}} = \Sigma_f^{\mathcal{D}} - \int_0^1 \Sigma_{xf,i}^{\mathcal{D}\top}(\tau)\Sigma_{x,i}^{\mathcal{D}-1}(\tau)\Sigma_{xf,i}^{\mathcal{D}}(\tau)d\tau$.

**Assumption 2.**

(i) $\lambda_{\min}(\Sigma_f^{\mathcal{D}})$, $\lambda_{\min}(\Sigma_{x,i}^{\mathcal{D}}(\tau))$, $\lambda_{\min}(\Omega_{f,i}^{\mathcal{D}}) > c$, almost surely for some positive constant $c$, and $\mathbb{E}(\|\Omega_{f,i}^{\mathcal{D}}\|_F) < \infty$, for any given $i$ and $\tau$.

(ii) $N^{-1}\sum_{i=1}^N \lambda_i^0\lambda_i^{0\top} = \Sigma_\lambda + O_P(N^{-1/2})$ and $\Sigma_\lambda$ is positive definite.

(iii) Random errors $\{\varepsilon_{it}\}$ satisfy

$$\sum_{i=1}^N\sum_{j=1}^N\sum_{t_1=1}^T\sum_{t_2=1}^T\sum_{t_3=1}^T\sum_{t_4=1}^T \mathbb{E}\left(|Cov_D(\varepsilon_{it_1}\varepsilon_{it_2}, \varepsilon_{jt_3}\varepsilon_{jt_4})|\right) \le CNT^2,$$

$$\sum_{i_1=1}^N\sum_{i_2=1}^N\sum_{i_3=1}^N\sum_{i_4=1}^N\sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\left(|Cov_D(\varepsilon_{i_1t}\varepsilon_{i_2t}, \varepsilon_{i_3s}\varepsilon_{i_4s})|\right) \le CN^2T.$$

(iv) For some $\delta^* \in (0,\delta)$, $NT^{-(4+\delta^*)/4} \to 0$. Additionally, $N^{\delta^\dagger}T^{-\theta}h^{-3-\theta}(\log T)^{1+2\theta} \to 0$, where $\delta^\dagger = \frac{6+\delta}{4+\delta} - \frac{2(1+\theta)}{2+\delta}$. Here $\theta$ and $\delta$ are defined in Assumption 1.

Assumption 2 imposes additional conditions on the non-singularity of covariance matrices, the convergence of $N^{-1}\sum_{i=1}^N \lambda_i^0\lambda_i^{0\top}$, and the weak cross-sectional

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

and serial dependence of random errors. The justification of Assumption 2 is available in Appendix A.1.

THEOREM 3.1. *Let Assumptions 1 and 2 hold. As $N, T \to \infty$,*

(1) $N^{-1/2} \left\| \widehat{B}^{(0)}(\tau) - B(\tau) \right\|_F = O_P\left( \max\left\{ h^{3/2}, \phi_{NT} \right\} h^{1/2} \right)$,
(2) $T^{-1/2} \left\| \widehat{F}^{(0)} - F^0 H \right\|_F = O_P\left( \max\left\{ h^{3/2}, \phi_{NT} \right\} h^{1/2} \right)$,
(3) $N^{-1/2} \left\| \widehat{\Lambda}^{(0)} - \Lambda^0 H^{-1\top} \right\|_F = O_P\left( \max\left\{ h^{3/2}, \phi_{NT} \right\} h^{1/2} \right)$,

*where $H = (NT)^{-1} \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\top} F^{0\top} \widehat{F}^{(0)} V_{NT}^{-1}$ is a rotation matrix.*

In Theorem 3.1, the rate of convergence is determined by two sources: nonparametric local linear approximation and nuclear-norm regularization estimation, with the latter producing a convergence rate of $O_P(\phi_{NT}\sqrt{h})$, which is slower than the optimal root-$Th$ rate in view of Assumption 1(v). Theorem 3.1(2) and (3) demonstrates the PCA estimators' consistency up to a rotation matrix $H$ as expected (see, for example, Bai, 2009).

Building on Theorem 3.1, the following theorem establishes the rates of convergence associated with the DLS algorithm, suggesting that the shrinkage estimation biases can be reduced through iterations.

THEOREM 3.2. *Let Assumptions 1 and 2 hold, and suppose $n \asymp \max\{\log N, \log T\}$. As $N, T \to \infty$,*

(1) $\left\| \widehat{f}_t^{(n)} - H^\top f_t^0 \right\| = O_P\left( \max\{h^2, N^{-1/2}, T^{-1}\} \right)$, *for each given t,*
(2) $\left\| \widehat{\lambda}_i^{(n)} - H^{-1} \lambda_i^0 \right\| = O_P\left( \max\{h^2, N^{-1}, T^{-1/2}\} \right)$, *for each given i,*
(3) $\left\| \widehat{\beta}_i^{(n)}(\tau) - \beta_i(\tau) \right\| = O_P\left( \max\{h^2, (Th)^{-1/2}\} \right)$, *for each given i and $\tau$.*

Theorem 3.2 demonstrates that our iteration algorithm can improve the estimation accuracy in the sense that the shrinkage bias in initial estimation is eliminated and the regularization parameter $\phi_{NT}$ plays no role in the rates of convergence, as the number of iterations diverges at an appropriate rate ($\max\{\log N, \log T\}$).

To establish central limit theorems (CLTs), additional conditions are required.

**Assumption 3.**

(i) $\Sigma_f^{\mathcal{D}} = \Sigma_f$, $\Omega_{f,i}^{\mathcal{D}} = \Omega_{f,i}$, and $\Sigma_{x,i}^{\mathcal{D}}(\tau) = \Sigma_{x,i}(\tau)$ almost surely, for each given $i$ and $\tau$.
(ii) $N^{-1/2} \sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it} \xrightarrow{D} \mathcal{N}(0, \Sigma_{\lambda\varepsilon}^0)$ and $\Sigma_{\lambda\varepsilon}^0$ is positive definite, for each given $t$.
(iii) $T^{-1/2} \sum_{t=1}^{T} z_{it} \varepsilon_{it} \xrightarrow{D} \mathcal{N}(0, \Sigma_{z\varepsilon,i}^0)$ and $\Sigma_{z\varepsilon,i}^0$ is positive definite, for each given $i$, where $z_{it} = f_t^0 - \Sigma_{xf,i}^{\mathcal{D}\top}(\tau_t) \Sigma_{x,i}^{\mathcal{D}-1}(\tau_t) x_{it}$.
(iv) $(Th)^{-1/2} \sum_{t=1}^{T} K_{t,0}(\tau) x_{it} \varepsilon_{it} \xrightarrow{D} \mathcal{N}\left(0, \Sigma_{x\varepsilon,i}^0(\tau)\right)$ and $\Sigma_{x\varepsilon,i}^0(\tau)$ is positive definite, for each given $i$ and $\tau$.

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

CLTs in Assumption 3 are routine requirements adopted in the literature (e.g., Assumption E of Bai, 2009). Let $V_{\lambda f}$ be an $r_0 \times r_0$ diagonal matrix that contains the eigenvalues of $\Sigma_\lambda^{1/2} \Sigma_f \Sigma_\lambda^{1/2}$ in descending order, and let $U_{\lambda f}$ be the corresponding orthogonal eigenvector matrix that satisfies $U_{\lambda f}^\top U_{\lambda f} = I_{r_0}$. Additionally, let $\mathcal{Q} = V_{\lambda f}^{1/2} U_{\lambda f}^\top \Sigma_\lambda^{-1/2}$ and $H_0 = \Sigma_\lambda \mathcal{Q}^\top V_{\lambda f}^{-1}$.

With Assumption 3, the following theorem summarizes the DLS estimators' asymptotic distributions.

THEOREM 3.3. *Let Assumptions 1–3 hold, and suppose $n \asymp \max\{\log N, \log T\}$. As $N, T \to \infty$,*

(1) *Additionally, if $N/T^2 \to 0$ and $Nh^4 \to 0$,*

$$\sqrt{N}\left(\widehat{f}_t^{(n)} - H^\top f_t^0\right) \xrightarrow{D} \mathcal{N}(0_{r_0}, \Sigma_{\lambda\varepsilon}),$$

*for each given t, where $\Sigma_{\lambda\varepsilon} = V_{\lambda f}^{-1} \mathcal{Q} \Sigma_{\lambda\varepsilon}^0 \mathcal{Q}^\top V_{\lambda f}^{-1}$.*

(2) *Additionally, if $T/N^2 \to 0$ and $Th^4 \to 0$,*

$$\sqrt{T}\left(\widehat{\lambda}_i^{(n)} - H^{-1}\lambda_i^0\right) \xrightarrow{D} \mathcal{N}(0_{r_0}, \Sigma_{z\varepsilon,i}),$$

*for each given i, where $\Sigma_{z\varepsilon,i} = H_0^{-1} \Omega_{f,i}^{-1} \Sigma_{z\varepsilon,i}^0 \Omega_{f,i}^{-1} H_0^{-1\top}$.*

(3) *For each given i and $\tau$,*

$$\sqrt{Th}\left(\widehat{\beta}_i^{(n)}(\tau) - \beta_i(\tau) - a_i(\tau)h^2\right) \xrightarrow{D} \mathcal{N}(0_p, \Sigma_{x\varepsilon,i}(\tau)),$$

*where $\Sigma_{x\varepsilon,i}(\tau) = \Sigma_{x,i}^{-1}(\tau)\Sigma_{x\varepsilon,i}^0(\tau)\Sigma_{x,i}^{-1}(\tau)$, $a_i(\tau) = \frac{\mu_2}{2}\beta_i''(\tau)(1+o(1))$, $\beta_i''(\tau)$ is the second-order derivative of $\beta_i(\tau)$, and $\mu_2 = \int u^2 K(u)du$.*

Theorem 3.3(1) and (2) establishes the asymptotic distributions of the factor and loading estimators, respectively, after imposing certain restrictions on the convergence rate of $h$ and the divergence rates of $N$ and $T$. As demonstrated in Theorem 3.3(3), the initial estimation biases are eliminated as a consequence of iterations, whereas a bias term $a_i(\tau)h^2$ remains due to the local linear approximation of time-varying functions. The presence of such a bias term is conventional in the literature on nonparametric time-varying models (e.g., Theorem 1 in Cai, 2007).

Typically, one might be interested in the following MG estimator (Pesaran, 2006):

$$\widehat{\beta}_w^{(n)}(\tau) = \sum_{i=1}^{N} w_{N,i} \widehat{\beta}_i^{(n)}(\tau),$$

where $w_{N,i}$ denotes the individual weight that satisfies $w_{N,i} > 0$ and $\sum_{i=1}^N w_{N,i} = 1$. In what follows, we follow Pesaran (2006) and employ the random coefficient assumption to establish the MG estimator's asymptotic distribution.

**Assumption 4.**

(i) $\beta_i(\tau) = \beta_0(\tau) + \epsilon_i(\tau)$, where $\beta_0(\tau)$ and $\epsilon_i(\tau)$ are $p$-dimensional vectors of unknown deterministic functions and time-varying random individual coefficients, respectively, for any $i$. Moreover, we assume $\epsilon_i(\tau) = \pi(\tau) \odot \varsigma_i$, where $\pi(\tau)$ is a $p$-dimensional vector of unknown deterministic functions capturing time-varying standard deviation of $\epsilon_i(\tau)$, and $\varsigma_i$ is a $p$-dimensional vector of random errors with zero mean and unit variance.

(ii) Unknown functions $\beta_0(\tau)$ and $\pi(\tau)$ have continuous derivatives of up to the second order on the support $\tau \in [0,1]$.

(iii) $\{\varsigma_i\}$ are independent of $\{v_{jt}, \lambda_j^0, f_t^0, \varepsilon_{jt}\}$ for any $(i,j,t)$. Moreover, $\varsigma_i$ is i.i.d. with $\mathbb{E}(\varsigma_i \varsigma_i^\top) = \Sigma_\varsigma$. Let $\Sigma_\epsilon(\tau) = \Sigma_\varsigma \odot (\pi(\tau)\pi(\tau)^\top)$ be a positive definite matrix, for $\tau \in [0,1]$.

Assumption 4 extends the model in Assumption 4 of Pesaran (2006) by allowing for unknown smooth time variations in the mean and variance of random coefficients.

THEOREM 3.4. *Let Assumptions 1–4 hold, and suppose $n \asymp \max\{\log N, \log T\}$. As $N, T \to \infty$,*

$$\sqrt{\gamma_{N,w}} \left( \widehat{\beta}_w^{(n)}(\tau) - \beta_0(\tau) - a_0(\tau) h^2 \right) \xrightarrow{D} \mathcal{N}(0_p, \Sigma_\epsilon(\tau)),$$

*for any given $\tau$, where $\gamma_{N,w} = \left( \sum_{i=1}^N w_{N,i}^2 \right)^{-1}$, $a_0(\tau) = \frac{\mu_2}{2} \beta_0''(\tau)(1 + o(1))$, and $\Sigma_\epsilon(\tau)$ is defined in Assumption 4.*

Up to this point, we have completed our investigation on the DLS algorithm. In Section 4, we address two practical issues: (1) testing time-invariant regression coefficients and (2) selecting the number of factors.

## 4. CONSTANCY TEST AND FACTOR NUMBER SELECTION

This section proposes a constancy test and an information criterion for factor number selection.

### 4.1. Constancy Test on Regression Coefficients

We are interested in testing the null hypothesis of time-constant parameters as follows:

$$\mathcal{H}_0: \quad \beta_i(\tau) = \beta_i^0, \text{ for } i = 1, \ldots, N,$$

where $\beta_i^0$ is a $p \times 1$ vector of unknown parameters. For power analysis, we consider the following local alternatives:

$$\mathcal{H}_1: \quad \beta_i(\tau) = \beta_i^0 + \nu_{NT} \Delta_{\beta,i}(\tau) \text{ for some } i,$$

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

where $\Delta_{\beta,i}(\tau)$ is a sequence of measurable and uniformly bounded nonparametric functions of $\tau$, and $\nu_{NT}$ satisfies $\nu_{NT} \to 0$ as $N,T \to \infty$. Let $N_a$ be the number of individuals that violate the time-constant null hypothesis.

To proceed, we define

$$e_{it} = y_{it} - x_{it}^\top \beta_i^0 - \lambda_i^{0\top} f_t^0. \qquad \textbf{(4.1)}$$

In view of (2.1) and (4.1), we have $e_{it} = \varepsilon_{it}$ under $\mathcal{H}_0$ and $e_{it} = \varepsilon_{it} + \nu_{NT} x_{it}^\top \Delta_{\beta,i}(\tau_t)$ for some $i$ under $\mathcal{H}_1$. Following Gao and Gijbels (2008) and Su et al. (2015), we construct a test statistic based on the restricted estimation residuals:

$$L_{NT} = \frac{1}{NT\sqrt{h}} \sum_{i=1}^{N} \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} K\left(\frac{\tau_t - \tau_s}{h}\right) \widehat{e}_{it} \widehat{e}_{ns},$$

where $\{\widehat{e}_{it}\}$ are the estimated residuals under the null hypothesis. For the sake of space, Appendix B.1 of the Supplementary Material provides a parametric version of the DLS algorithm in detail, which is used to calculate $\{\widehat{e}_{it}\}$ here. Denote $\widetilde{\beta}_i^{(n)}$, $\widetilde{f}_t^{(n)}$, and $\widetilde{\lambda}_i^{(n)}$ as the parametric iterative estimators of $\beta_i^0, f_t^0$, and $\lambda_i^0$, respectively. Then, we compute the restricted residuals: $\widehat{e}_{it} = y_{it} - x_{it}^\top \widetilde{\beta}_i^{(n)} - \widetilde{\lambda}_i^{(n)\top} \widetilde{f}_t^{(n)}$. Some additional conditions are introduced in the following assumption.

**Assumption 5.**

(i) $\{\varepsilon_t\}$ are martingale difference sequences (m.d.s.) adapted to the filtration $\{\mathcal{F}_t\}$, where $\mathcal{F}_t$ is the sigma-field generated by $\{\mathcal{D}, \varepsilon_t, \varepsilon_{t-1}, \ldots\}$, and $\varepsilon_t$ satisfies $\mathbb{E}(\varepsilon_{it}\varepsilon_{jt} | \mathcal{F}_{t-1}) = \mathbb{E}(\varepsilon_{it}\varepsilon_{jt}) = \sigma^2_{\varepsilon,ij}$ (a.s.). Moreover, $\overline{\sigma}^2_\varepsilon = \lim_{N\to\infty} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sigma^2_{\varepsilon,ij}$ is a positive and finite constant.

(ii) Assume that $\max_{1\le t<s} \mathbb{E}\left(\left|\sum_{i=1}^N \sum_{j=1}^N \varepsilon_{it}\varepsilon_{js}\right|^4\right) \le CN^4$, for $s = 2,\ldots,T$.

(iii) Assume that $T^{2-\frac{\theta\delta}{2(4+\delta)}} h^{2+\frac{\theta\delta}{2(4+\delta)}} \log T^{-\frac{1}{2}} \to 0$, where $\theta$ and $\delta$ are defined in Assumption 1.

The justifications of these conditions are provided in Appendix A.1. Under Assumption 5, simple algebra shows that $L_{NT}$'s asymptotic covariance is $\sigma_L^2 = 2\nu_0 \overline{\sigma}_\varepsilon^4$, where $\nu_0 = \int K^2(v)dv$. Therefore, the final version of the test statistic is as follows:

$$\breve{L}_{NT} = \frac{1}{\sqrt{\widehat{\sigma}_L^2}} L_{NT},$$

where $\widehat{\sigma}_L^2 = 2\nu_0 \widehat{\overline{\sigma}}_\varepsilon^4$ and $\widehat{\overline{\sigma}}_\varepsilon^2 = (NT)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \widehat{e}_{it}\widehat{e}_{jt}$.

The asymptotic properties of $\breve{L}_{NT}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ are studied in the following theorem.

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

THEOREM 4.1. *Let Assumptions 1–3 and 5 hold. As $N,T \to \infty$ simultaneously,*

(1) $\breve{L}_{NT} \xrightarrow{D} \mathcal{N}(0,1)$ *under* $\mathcal{H}_0$.
(2) *If, in addition,* $N_a$ *and* $\nu_{NT}$ *satisfy* $\nu_{NT} \to 0$ *and* $N_a N^{-\frac{1}{2}} T^{\frac{1}{2}} h^{\frac{1}{4}} \nu_{NT} \to \infty$, *then* $P(\breve{L}_{NT} > L_\alpha) \to 1$ *under* $\mathcal{H}_1$, *where* $L_\alpha$ *denotes the* $\alpha$*-level critical value of* $\mathcal{N}(0,1)$.

Theorem 4.1(1) establishes the asymptotic normality of the standardized test statistic under the null hypothesis, whereas Theorem 4.1(2) shows that the test is asymptotically consistent under a sequence of local alternatives.

By Theorem 4.1, one can use the theoretical critical values based on the distribution $\mathcal{N}(0,1)$. However, in practice, one may turn to the bootstrap method to achieve better finite-sample performance (e.g., Gao and Gijbels, 2008; Su et al., 2015). Therefore, we further follow Su and Wang (2017) and construct the critical values using a dependent bootstrap method. Let $\widetilde{\Sigma}_\varepsilon$ be a shrinkage version of the covariance estimator, and let its $(i,j)$th element be $\widetilde{\sigma}^2_{\varepsilon,ij} = \widehat{\sigma}^2_{\varepsilon,ij}(1-\epsilon)^{|j-i|}$ for each $i$ and $j$, where $\widehat{\sigma}^2_{\varepsilon,ij} = T^{-1}\sum_{t=1}^{T}\widehat{e}_{it}\widehat{e}_{jt}$ and $\epsilon$ is a pre-specified small positive number that ensures the positive definiteness of $\widetilde{\Sigma}_\varepsilon$. The bootstrap procedure is summarized as follows.

*Step 1.* Obtain parametric iterative estimators $\widetilde{\beta}_i^{(n)}, \widetilde{f}_t^{(n)}$, and $\widetilde{\lambda}_i^{(n)}$ under $\mathcal{H}_0$, and compute restricted residuals:

$$\widehat{e}_{it} = y_{it} - x_{it}^\top \widetilde{\beta}_i^{(n)} - \widetilde{\lambda}_i^{(n)\top}\widetilde{f}_t^{(n)}.$$

With $\{\widehat{e}_{it}\}$, the test statistic $\breve{L}_{NT}$ and covariance estimator $\widetilde{\Sigma}_\varepsilon$ can be constructed.

*Step 2.* Compute the bootstrap error terms $(e^*_{1t},\ldots,e^*_{Nt})^\top = \widetilde{\Sigma}_\varepsilon^{1/2}\eta_t$, where $\eta_t = (\eta_{1t},\ldots,\eta_{Nt})^\top$ follow i.i.d. $\mathcal{N}(0,1)$, and construct $\{y^*_{it}\}$ as

$$y^*_{it} = x_{it}^\top \widetilde{\beta}_i^{(n)} + \widetilde{\lambda}_i^{(n)\top}\widetilde{f}_t^{(n)} + e^*_{it}.$$

Thereafter, the bootstrap sample is given as $\{y^*_{it}, x_{it}\}$.

*Step 3.* Obtain bootstrap estimators $\widetilde{\beta}_i^*, \widetilde{f}_t^*$, and $\widetilde{\lambda}_i^*$ under $\mathcal{H}_0$, and compute the restricted residuals:

$$\widehat{e}^*_{it} = y^*_{it} - x_{it}^\top \widetilde{\beta}_i^* - \widetilde{\lambda}_i^{*\top}\widetilde{f}_t^*.$$

The bootstrap test statistic $\widetilde{L}^*_{NT}$ can be constructed using $\{\widehat{e}^*_{it}\}$.

*Step 4.* Repeat Steps 2 and 3 $M$ times and obtain the bootstrap distribution from $\{\widetilde{L}^*_{NT,m}\}_{m=1}^M$. The bootstrap $\alpha$-level critical $L^*_\alpha$ is calculated as $P^*(\widetilde{L}^*_{NT,m} \geq L^*_\alpha) = \alpha$, where $P^*$ denotes the probability measure conditional on the observed sample $\{y_{it}, x_{it}\}$.

The asymptotic properties of $\widetilde{L}^*_{NT}$ are summarized in the following theorem.

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

THEOREM 4.2. *Let Assumptions 1–3 and 5 hold. As $N,T\to\infty$ simultaneously, $\widetilde{L}^*_{NT}\xrightarrow{D^*}\mathcal{N}(0,1)$ in probability, where $D^*$ denotes the convergence in distribution with respect to $P^*$ conditional on the original sample.*

Theorem 4.2 establishes the asymptotic distribution of $\widetilde{L}^*_{NT}$, indicating that the bootstrap test statistic exhibits the same asymptotic behavior as $\breve{L}_{NT}$ conditional on the original sample.

## 4.2. Factor Number Selection

This section proposes an information criterion method to determine the number of factors ($r_0$) by adding a penalty term to the log value of the mean-squared initial estimation residuals. Recall that we can obtain the consistent initial estimator $\widehat{B}^{(0)}(\tau)$ without the need to know the true value of $r_0$. Let $\widehat{F}^{(0,r)}$ and $\widehat{\Lambda}^{(0,r)}$ be the initial factor and loading estimators, respectively, with $r>1$ in particular.

We construct the following information criterion:

$$\mathrm{IC}(r)=\log\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-x_{it}^{\top}\widehat{\beta}_i^{(0)}(\tau_t)-\widehat{\lambda}_i^{(0,r)\top}\widehat{f}_t^{(0,r)}\right)^2\right)+d_{NT}\cdot r, \tag{4.2}$$

where $d_{NT}$ is a penalty term satisfying certain restrictions to be specified later. For the case $r=0$, let $\mathrm{IC}(0)=\log(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}^{\top}\widehat{\beta}_i^{(0)}(\tau_t))^2)$. For a prespecified large integer $r_{\max}$, we estimate $r_0$ by

$$\widehat{r}=\underset{0\le r\le r_{\max}}{\operatorname{argmin}}\,\mathrm{IC}(r). \tag{4.3}$$

The following theorem establishes the consistency.

THEOREM 4.3. *Let Assumptions 1 and 2 hold. Additionally, if $d_{NT}$ satisfies (i) $d_{NT}\to 0$ and (ii) $\frac{d_{NT}}{\max\left\{h^3,\phi_{NT}^2\right\}h}\to\infty$, as $N,T\to\infty$, $Pr(\widehat{r}=r_0)\to 1$.*

Theorem 4.3 reveals that the number of factors can be consistently estimated using the information criterion method defined in (4.2). The finite-sample performance of this method is assessed using simulation examples in Section 5. An alternative method for determining $r_0$ is the singular-value thresholding approach that utilizes the low-rank structure of $\Gamma^0=\Lambda^0F^{0\top}$. Miao, Phillips, and Su (2023) provide a complete description of this method.

## 5. SIMULATION STUDIES

Simulation experiments are performed in this section to assess the finite-sample performance of DLS estimators, the constancy test procedure, and the information criterion for factor number selection.

### 5.1. DLS Estimation

This section presents three simulation examples: Examples 5.1 and 5.2 employ heterogeneous time-varying regression coefficients, and Example 5.3 considers a homogeneous time-varying model.

**Example 5.1** Consider the following data generating process (DGP):

$$Y_{it} = x_{it,1}\beta_{1i}(\tau_t) + x_{it,2}\beta_{2i}(\tau_t) + \lambda_{1i}^0 f_{1t}^0 + \lambda_{2i}^0 f_{2t}^0 + \varepsilon_{it}, \tag{5.1}$$

where $(x_{it,1}, x_{it,2}) = (\lambda_{1i}^0 f_{1t}^0 + \lambda_{2i}^0 f_{2t}^0 + \eta_{it,1}, \eta_{it,2})$, $(\beta_{1i}(\tau), \beta_{2i}(\tau)) = (\sin(\pi\tau) + \epsilon_{1i}, \cos(\pi\tau) + \epsilon_{2i})$ with $\eta_{it,1}$, $\eta_{it,2}$, $\epsilon_{1i}$, and $\epsilon_{2i}$ being generated from i.i.d. $\mathcal{N}(0,1)$, and the factors $(f_{1t}^0, f_{2t}^0)$ are both AR(1) processes: $(f_{1t}^0, f_{2t}^0) = (\rho_{f_1} f_{1,t-1}^0 + v_{f_1,t}, \rho_{f_2} f_{2,t-1}^0 + v_{f_2,t})$. Here, $\rho_{f_1} = 0.6$ and $\rho_{f_2} = 0.4$, and $v_{f_1,t}$ and $v_{f_2,t}$ follow i.i.d. $\mathcal{N}(0,1)$. Let $(\lambda_{11}^0, \ldots, \lambda_{1N}^0)^\top$, $(\lambda_{21}^0, \ldots, \lambda_{2N}^0)^\top$, and $(\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})^\top$ be generated as $N$-dimensional vectors of independent Gaussian variables with zero mean and the $(i,j)$th element of the covariance matrix as $\sigma_{ij} = 0.5^{|i-j|}$, for $i,j = 1,2,\ldots,N$.

**Example 5.2** Consider the DGP (5.1), where $(x_{it,1}, x_{it,2}) = (\sin(\pi\tau_t) + \lambda_{1i}^0 f_{1t}^0 + \lambda_{2i}^0 f_{2t}^0 + \eta_{it,1}, \tau_t + \eta_{it,2})$, $(\beta_{1i}(\tau), \beta_{2i}(\tau)) = (1 + \tau + \epsilon_{1i}, 1 + \tau^3 + \epsilon_{2i})$ and the other variables are generated in the same manner as in Example 5.1.

**Example 5.3** Consider the following DGP:

$$Y_{it} = x_{it,1}\beta_1(\tau_t) + x_{it,2}\beta_2(\tau_t) + \lambda_{1i}^0 f_{1t}^0 + \lambda_{2i}^0 f_{2t}^0 + \varepsilon_{it},$$

where $(\beta_1(\tau), \beta_2(\tau)) = (\sin(\pi\tau), \cos(\pi\tau))$ and the other variables are generated in the same manner as in Example 5.1.

The number of factors is regarded as known in this section. We select the regularization parameter as $\phi_{NT} = (NT)^{-1/10}$. The Epanechnikov kernel function $K(\mu) = \frac{3}{4}(1-\mu^2)I(|\mu| \le 1)$ is used in the local linear estimation. The leave-one-out cross-validation method is employed to select the optimal bandwidth. Specifically, $h_{cv}$ is selected to minimize the following objective function:

$$h_{cv} = \min_h \sum_{i=1}^N \sum_{t=1}^T \left( Y_{it} - x_{it,1}\widehat{\beta}_{1i}^{(-t)}(\tau_t) - x_{it,2}\widehat{\beta}_{2i}^{(-t)}(\tau_t) \right)^2,$$

where $\left(\widehat{\beta}_{1i}^{(-t)}(\tau), \widehat{\beta}_{2i}^{(-t)}(\tau)\right)$ are the leave-one-out estimates.

After 1,000 replications ($R = 1{,}000$), we calculate the average of the mean-squared errors (AMSEs) of the initial and iterative DLS estimates:

$$\text{AMSE}_{\beta,0} = \frac{1}{2NTR} \sum_{k=1}^R \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^2 \left( \widehat{\beta}_{mi}^{(0,k)}(\tau_t) - \beta_{mi}(\tau_t) \right)^2,$$

$$\text{AMSE}_{\beta,n} = \frac{1}{2NTR} \sum_{k=1}^R \sum_{i=1}^N \sum_{t=1}^T \sum_{m=1}^2 \left( \widehat{\beta}_{mi}^{(n,k)}(\tau_t) - \beta_{mi}(\tau_t) \right)^2,$$

where $\widehat{\beta}_{mi}^{(0,k)}(\tau)$ and $\widehat{\beta}_{mi}^{(n,k)}(\tau)$ denote the initial and DLS estimates of $\beta_{mi}(\tau)$, respectively, in the $k$th replication. To assess the estimation accuracy of factors and loadings, we compute their second canonical-correlation coefficients (SCCs) in each replication.[1] Table 1 reports AMSEs and averaged SCCs for Examples 5.1–5.3.

As presented in Table 1, the AMSEs decrease as $T$ diverges in all three examples. Additionally, it is unsurprising that the AMSEs of the iterative estimates decrease more rapidly than those of initial estimates. These simulation results are in line with the established asymptotic theory which states that both initial and DLS estimators are consistent irrespective of the correlation between regressors and common factors, and iterations can substantially reduce the initial estimation bias. In terms of SCCs, they all tend to increase toward one with growing $N$ and $T$. This finding provides some numerical evidence for the loading and factor estimators' consistency up to a rotation matrix. The values of iterative loading and factor estimates are invariably greater than those for initial estimates, also demonstrating the improved estimation accuracy of the DLS iterative procedure. The results of Example 5.2 confirm the robustness of our methodology to nonparametric deterministic trends in regressors. In view of the simulation results for Examples 5.1 and 5.3, wherein the difference in AMSEs and SCCs is not evident, the proposed estimation approach can be applied to estimate both homogeneous and heterogeneous models. For Example 5.3, we construct MG estimates with $w_{N,i} = 1/N$ and calculate their simulated 95% confidence intervals, which are reported in Figure 1. The bandwidth $h = T^{-1/3}$ is used for bias reduction.

## 5.2. Constancy Test

This section utilizes two simulation examples to demonstrate the constancy test's size and power performance.

**Example 5.4** Consider the following DGP:

$$Y_{it} = x_{it,1}\beta_{1i} + x_{it,2}\beta_{2i} + \lambda_{1i}^0 f_{1t}^0 + \lambda_{2i}^0 f_{2t}^0 + \varepsilon_{it},$$

where $(\beta_{1i}, \beta_{2i}) = (1+\epsilon_{1i}, 1+\epsilon_{2i})$, $\epsilon_{1i}$ and $\epsilon_{2i}$ follow i.i.d. $\mathcal{N}(0,1)$, and the other variables are generated in the same manner as in Example 5.2.

**Example 5.5** Consider the following DGP:

$$Y_{it} = x_{it,1}\beta_{1i}(\tau_t) + x_{it,2}\beta_{2i}(\tau_t) + \lambda_{1i}^0 f_{1t}^0 + \lambda_{2i}^0 f_{2t}^0 + \varepsilon_{it},$$

[1] The canonical-correlation coefficient, which is robust to the rotation matrix, is widely used in the literature on factor models (e.g., Bai and Li, 2012) as a tool to measure estimation accuracy. We briefly introduce its definition. The first canonical-correlation coefficient of $\widehat{f}_t^{(n)}$ and $f_t^0$ is defined as the maximum value of correlation coefficients $\mathrm{corr}(a^\top \widehat{f}_t^{(n)}, b^\top f_t^0)$ for all $r_0 \times 1$ vectors $a$ and $b$. The SCC is defined as the maximum value of correlation coefficients $\mathrm{corr}(c^\top \widehat{f}_t^{(n)}, d^\top f_t^0)$ for all $r_0 \times 1$ vectors $c$ and $d$ that are orthogonal to $a$ and $b$. Therefore, the SCC increasing toward one indicates the consistency of factor and loading estimates (up to a rotation matrix).

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

**TABLE 1.** AMSEs and SCCs for Examples 5.1–5.3

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Example 5.1 | | | | | | | | | | | | |
| | $AMSE_{\beta,0}$ | | | | $SCC_{\lambda,0}$ | | | | $SCC_{f,0}$ | | | |
| $N/T$ | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 |
| 20 | 0.4781 | 0.4133 | 0.3544 | 0.3326 | 0.3707 | 0.5656 | 0.7789 | 0.9099 | 0.2330 | 0.3006 | 0.4205 | 0.5353 |
| 40 | 0.4736 | 0.3848 | 0.3294 | 0.3019 | 0.4796 | 0.7175 | 0.8973 | 0.9383 | 0.2843 | 0.4680 | 0.6143 | 0.6659 |
| 80 | 0.4579 | 0.3653 | 0.3058 | 0.2833 | 0.5872 | 0.8472 | 0.9271 | 0.9510 | 0.4063 | 0.5856 | 0.6877 | 0.7001 |
| | $AMSE_{\beta,n}$ | | | | $SCC_{\lambda,n}$ | | | | $SCC_{f,n}$ | | | |
| $N/T$ | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 |
| 20 | 0.4487 | 0.1385 | 0.0631 | 0.0597 | 0.4743 | 0.7681 | 0.9333 | 0.9715 | 0.5439 | 0.7659 | 0.8854 | 0.9156 |
| 40 | 0.3398 | 0.1157 | 0.0543 | 0.0529 | 0.7348 | 0.9005 | 0.9777 | 0.9863 | 0.7693 | 0.9084 | 0.9590 | 0.9716 |
| 80 | 0.3119 | 0.1105 | 0.0536 | 0.0504 | 0.8196 | 0.9585 | 0.9817 | 0.9920 | 0.8673 | 0.9613 | 0.9729 | 0.9893 |
| Example 5.2 | | | | | | | | | | | | |
| | $AMSE_{\beta,0}$ | | | | $SCC_{\lambda,0}$ | | | | $SCC_{f,0}$ | | | |
| $N/T$ | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 |
| 20 | 0.3831 | 0.2015 | 0.1735 | 0.1352 | 0.4454 | 0.6001 | 0.7374 | 0.8862 | 0.3439 | 0.3910 | 0.4537 | 0.5598 |
| 40 | 0.2648 | 0.1884 | 0.1581 | 0.1249 | 0.5217 | 0.7265 | 0.8911 | 0.9487 | 0.4580 | 0.5596 | 0.7208 | 0.7626 |
| 80 | 0.2897 | 0.1714 | 0.1412 | 0.1157 | 0.6241 | 0.8455 | 0.9270 | 0.9608 | 0.5745 | 0.7724 | 0.8522 | 0.8931 |
| | $AMSE_{\beta,n}$ | | | | $SCC_{\lambda,n}$ | | | | $SCC_{f,n}$ | | | |
| $N/T$ | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 |
| 20 | 0.2537 | 0.1376 | 0.0408 | 0.0278 | 0.4694 | 0.7025 | 0.8188 | 0.9692 | 0.5108 | 0.6979 | 0.7864 | 0.8750 |
| 40 | 0.2382 | 0.0976 | 0.0364 | 0.0246 | 0.6674 | 0.8435 | 0.9662 | 0.9884 | 0.7640 | 0.8734 | 0.9640 | 0.9809 |
| 80 | 0.2368 | 0.0778 | 0.0318 | 0.0219 | 0.7587 | 0.9467 | 0.9850 | 0.9952 | 0.8649 | 0.9709 | 0.9901 | 0.9929 |
| Example 5.3 | | | | | | | | | | | | |
| | $AMSE_{\beta,0}$ | | | | $SCC_{\lambda,0}$ | | | | $SCC_{f,0}$ | | | |
| $N/T$ | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 |
| 20 | 0.4905 | 0.4101 | 0.3504 | 0.3345 | 0.4050 | 0.5990 | 0.7826 | 0.9222 | 0.2215 | 0.3261 | 0.4360 | 0.5400 |
| 40 | 0.4741 | 0.3855 | 0.3290 | 0.3022 | 0.4771 | 0.7436 | 0.9040 | 0.9404 | 0.3223 | 0.4785 | 0.6133 | 0.6519 |
| 80 | 0.4584 | 0.3676 | 0.3058 | 0.2745 | 0.5768 | 0.8542 | 0.9240 | 0.9447 | 0.4157 | 0.6044 | 0.6898 | 0.7323 |
| | $AMSE_{\beta,n}$ | | | | $SCC_{\lambda,n}$ | | | | $SCC_{f,n}$ | | | |
| $N/T$ | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 | 20 | 40 | 80 | 120 |
| 20 | 0.5920 | 0.1283 | 0.0615 | 0.0503 | 0.5444 | 0.8254 | 0.9356 | 0.9870 | 0.5442 | 0.7950 | 0.8995 | 0.9224 |
| 40 | 0.3507 | 0.1136 | 0.0555 | 0.0467 | 0.6618 | 0.9128 | 0.9796 | 0.9885 | 0.7412 | 0.9170 | 0.9603 | 0.9732 |
| 80 | 0.3167 | 0.1085 | 0.0526 | 0.0437 | 0.8093 | 0.9580 | 0.9820 | 0.9915 | 0.8680 | 0.9626 | 0.9739 | 0.9838 |

where $(\beta_{1i}(\tau), \beta_{2i}(\tau)) = (1+\epsilon_{1i}, 1+\epsilon_{2i}) + \nu_{NT}(\sin(\pi\tau), \cos(\pi\tau))$, $\epsilon_{1i}$ and $\epsilon_{2i}$ follow i.i.d. $\mathcal{N}(0,1)$, and the other variables are generated in the same manner as in Example 5.2. In this example, as $N_a = N$, we use $\nu_{NT} = \log(NT)^{1/2}(NT)^{-1/2}h^{-1/4}$ to calculate the local power.

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

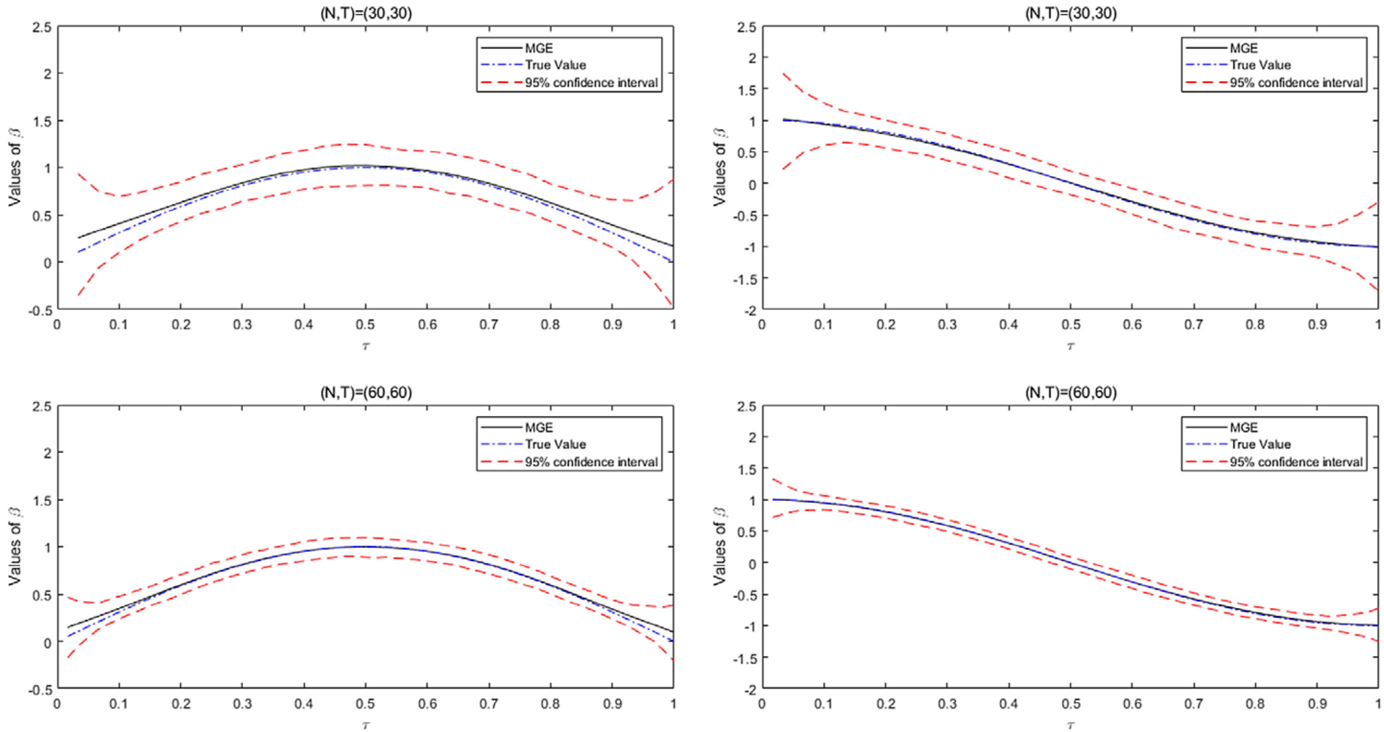


**FIGURE 1.** Confidence intervals of MG estimates for Example 5.3.

**TABLE 2.** Rejection rates for Examples 5.4 and 5.5

| | Example 5.4 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha = 0.01$ | | | $\alpha = 0.05$ | | | $\alpha = 0.10$ | | |
| $N/T$ | 20 | 40 | 80 | 20 | 40 | 80 | 20 | 40 | 80 |
| 20 | 0.005 | 0.012 | 0.006 | 0.038 | 0.063 | 0.055 | 0.096 | 0.089 | 0.104 |
| 40 | 0.016 | 0.007 | 0.013 | 0.061 | 0.040 | 0.047 | 0.113 | 0.108 | 0.093 |
| 80 | 0.014 | 0.010 | 0.011 | 0.060 | 0.046 | 0.052 | 0.112 | 0.100 | 0.103 |
| | Example 5.5 | | | | | | | | |
| | $\alpha = 0.01$ | | | $\alpha = 0.05$ | | | $\alpha = 0.10$ | | |
| $N/T$ | 20 | 40 | 80 | 20 | 40 | 80 | 20 | 40 | 80 |
| 20 | 0.055 | 0.164 | 0.512 | 0.141 | 0.413 | 0.750 | 0.249 | 0.595 | 0.848 |
| 40 | 0.081 | 0.423 | 0.751 | 0.240 | 0.772 | 0.914 | 0.286 | 0.835 | 0.960 |
| 80 | 0.102 | 0.709 | 0.890 | 0.311 | 0.893 | 0.976 | 0.374 | 0.947 | 0.993 |

For Examples 5.4 and 5.5, we test the following null hypothesis:

$$\mathcal{H}_0 : \beta_{1i}(\tau) = \beta_{1i}^0, \quad \beta_{2i}(\tau) = \beta_{2i}^0,$$

where $\beta_{1i}^0$ and $\beta_{2i}^0$ are unknown time-constant regression coefficients. The number of bootstraps to select the critical values is 500. After 1,000 replications, we compute the rejection rates under 1%, 5%, and 10% confidence levels. The simulation results of Examples 5.4 and 5.5 are reported in Table 2.

The null hypothesis $\mathcal{H}_0$ holds in Example 5.4. Therefore, the rejection rates in this example are simulated sizes. As presented in Table 2, simulated sizes are close to the corresponding significance levels across all the sample sizes. In addition, the rejection rates in Example 5.5, which generally increase toward one as $N$ and $T$ grow from 20 to 80, are simulated local powers, because the null hypothesis of time-constant coefficients does not hold in this example. In summary, the proposed test behaves reasonably well for both simulated sizes and powers.

## 5.3. Factor Number Selection

This section assesses the finite-sample performance of the proposed information criterion method in two additional simulation examples.

**Example 5.6** Consider the following DGP:

$$Y_{it} = x_{it,1}\beta_{1i}(\tau_t) + x_{it,2}\beta_{2i}(\tau_t) + \lambda_{1i}^0 f_{1t}^0 + \lambda_{2i}^0 f_{2t}^0 + \lambda_{3i}^0 f_{3t}^0 + \varepsilon_{it},$$

where $f_{3t}^0$ follows i.i.d. $\mathcal{N}(0,1)$, $(x_{1t,1}, \ldots, x_{Nt,1})^\top$, $(x_{1t,2}, \ldots, x_{Nt,2})^\top$, and $(\lambda_{31}^0, \ldots, \lambda_{3N}^0)^\top$ are generated as $N$-dimensional vectors of independent Gaussian variables with zero mean and covariance matrix with $(i,j)$th element: $\sigma_{ij} = 0.5^{|i-j|}$. The other variables are generated in the same manner as in Example 5.1.

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

**Table 3.** Correct-selection rates for Examples 5.6 and 5.7

| | Example 5.6 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Under-selection | | | Correct-selection | | | Over-selection | | |
| $N/T$ | 20 | 40 | 80 | 20 | 40 | 80 | 20 | 40 | 80 |
| 20 | 0.123 | 0.034 | 0 | 0.693 | 0.934 | 1 | 0.184 | 0.032 | 0 |
| 40 | 0.057 | 0 | 0 | 0.943 | 1 | 1 | 0 | 0 | 0 |
| 80 | 0.019 | 0 | 0 | 0.981 | 1 | 1 | 0 | 0 | 0 |
| | Example 5.7 | | | | | | | | |
| | Under-selection | | | Correct-selection | | | Over-selection | | |
| $N/T$ | 20 | 40 | 80 | 20 | 40 | 80 | 20 | 40 | 80 |
| 20 | 0.334 | 0.323 | 0.148 | 0.354 | 0.604 | 0.781 | 0.312 | 0.073 | 0.071 |
| 40 | 0.271 | 0.130 | 0.034 | 0.679 | 0.858 | 0.964 | 0.050 | 0.012 | 0.002 |
| 80 | 0.081 | 0.020 | 0 | 0.838 | 0.966 | 1 | 0.081 | 0.014 | 0 |

**Example 5.7** Consider the following DGP:

$$Y_{it} = x_{it,1}\beta_{1i}(\tau_t) + x_{it,2}\beta_{2i}(\tau_t) + \lambda_{1i}^0 f_{1t}^0 + \lambda_{2i}^0 f_{2t}^0 + \lambda_{3i}^0 f_{3t}^0 + \varepsilon_{it},$$

where $(x_{it,1}, x_{it,2}) = ((0.2\lambda_{1i}^0 + \gamma_{1i}^0)f_{1t}^0 + (0.2\lambda_{2i}^0 + \gamma_{2i}^0)f_{2t}^0 + \eta_{it,1}, \eta_{it,2})$ with $(\gamma_{11}^0, \ldots, \gamma_{1N}^0)^\top$ and $(\gamma_{21}^0, \ldots, \gamma_{2N}^0)^\top$ being generated as $N$-dimensional vectors of independent Gaussian variables with zero mean and the $(i,j)$th element of the covariance matrix as $\sigma_{ij} = 0.5^{|i-j|}$. The other variables are generated in the same manner as in Example 5.6.

Example 5.6 employs a DGP in which the regressors and factors are generated independently, and the interaction between such variables is allowed in Example 5.7. For both examples, $d_{NT} = (NT)^{-1/5}$ is used to construct the information criterion. After 1,000 replications, we calculate the rates of under-selection, correct-selection, and over-selection, and report the simulation results in Table 3. The correct-selection rates increase rapidly as $N$ and $T$ diverge in both examples. By contrast, the under-selection and over-selection rates decline significantly. In view of these simulation results, the proposed information criterion method performs reasonably effectively irrespective of the interaction between regressors and factors. However, this method exhibits significantly better finite-sample performance for uncorrelated regressors and factors when $N$ and $T$ are small (e.g., $N = T = 20$).

## 6. AN EMPIRICAL APPLICATION IN FINANCE

The determinants of fund performance have been extensively studied in the finance literature (e.g., Ferson and Schadt, 1996; Mamaysky, Spiegel, and Zhang, 2008; Blake et al., 2014). We examine the effects of three fund characteristic variables (namely, fund size, fund family size, and fund management fees) on mutual fund returns and allow time variations and heterogeneity in their associations.

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

**TABLE 4.** Moon and Perron's (2004) unit-root test results for regressor residuals

| Residuals in regressors | $t_a^*$ | $p$-value for $t_a^*$ | $t_b^*$ | $p$-value for $t_b^*$ |
|---|---|---|---|---|
| $\text{FFS}_{it}$ | −34.58 | < 0.01 | −13.76 | < 0.01 |
| $\text{MFS}_{it}$ | −33.06 | < 0.01 | −14.5 | < 0.01 |
| $\text{MC}_{it}$ | −37.33 | < 0.01 | −12.88 | < 0.01 |

**TABLE 5.** Values of IC($r$)

| $r$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $IC(r)$ | −6.1126 | −8.1058 | −8.2266 | −8.2358 | −8.2347 | −8.2292 | −8.2000 | −8.1472 | −8.0975 | −8.0278 | −7.9514 |

We collect the monthly return data of U.S. mutual funds from the Center for Research in Security Prices (CRSP) Survivor-Bias-Free Mutual Fund database for the 2008–2019 period ($T = 144$). After excluding funds with incomplete records, initial total net assets (TNAs) under 10 million dollars, or less than 90% of portfolio holdings invested in equity markets, we obtain a sample with 117 mutual funds ($N = 117$). In what follows, we introduce the dependent and explanatory variables of interest. Mutual fund excess returns ($\text{ER}_{it}$), as the dependent variable, are calculated as the difference between the fund returns and risk-free rates, which are also collected from CRSP. The fund family size ($\text{FFS}_{it}$) is defined as the ratio of fund family TNAs under management to the average value of all fund family TNAs in the previous month. The mutual fund size ($\text{MFS}_{it}$) is defined as the ratio of the fund's TNAs under management to the average value of TNAs across all funds in the previous month, and the management charge of the fund family ($\text{MC}_{it}$) is defined as the average value of charges for the funds under management in the current month. We employ the following panel data model with time-varying coefficients, heterogeneity, and IFEs:

$$\text{ER}_{it} = \beta_{0i}(\tau_t) + \text{FFS}_{it}\beta_{1i}(\tau_t) + \text{MFS}_{it}\beta_{2i}(\tau_t) + \text{MC}_{it}\beta_{3i}(\tau_t) + \lambda_i^{0\top} f_t^0 + \varepsilon_{it}, \qquad \textbf{(6.1)}$$

where the number of factors needs to be determined.

As described in Section 3, the proposed methodology is robust to unknown deterministic trends in regressors. To demonstrate its validity in this application, we first estimate the trends ($g_i(\tau)$) of regressors using the local linear method and thereafter examine the stationarity of residuals. Moon and Perron's (2004) method, as a second-generation panel data unit-root test that allows for cross-sectional dependence, is employed to detect the residuals' nonstationary behavior. Specifically, we compute the feasible $t$-statistics $t_a^*$ and $t_b^*$ defined in Lemma 4 of Moon and Perron (2004). As presented in Table 4, the test results reveal that the null hypothesis of unit root is rejected for all three explanatory variables after we remove the nonparametric deterministic trends.

We use the Epanechnikov kernel function to construct estimates, and the leave-one-out cross-validation method is adopted to select the optimal bandwidth. For

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

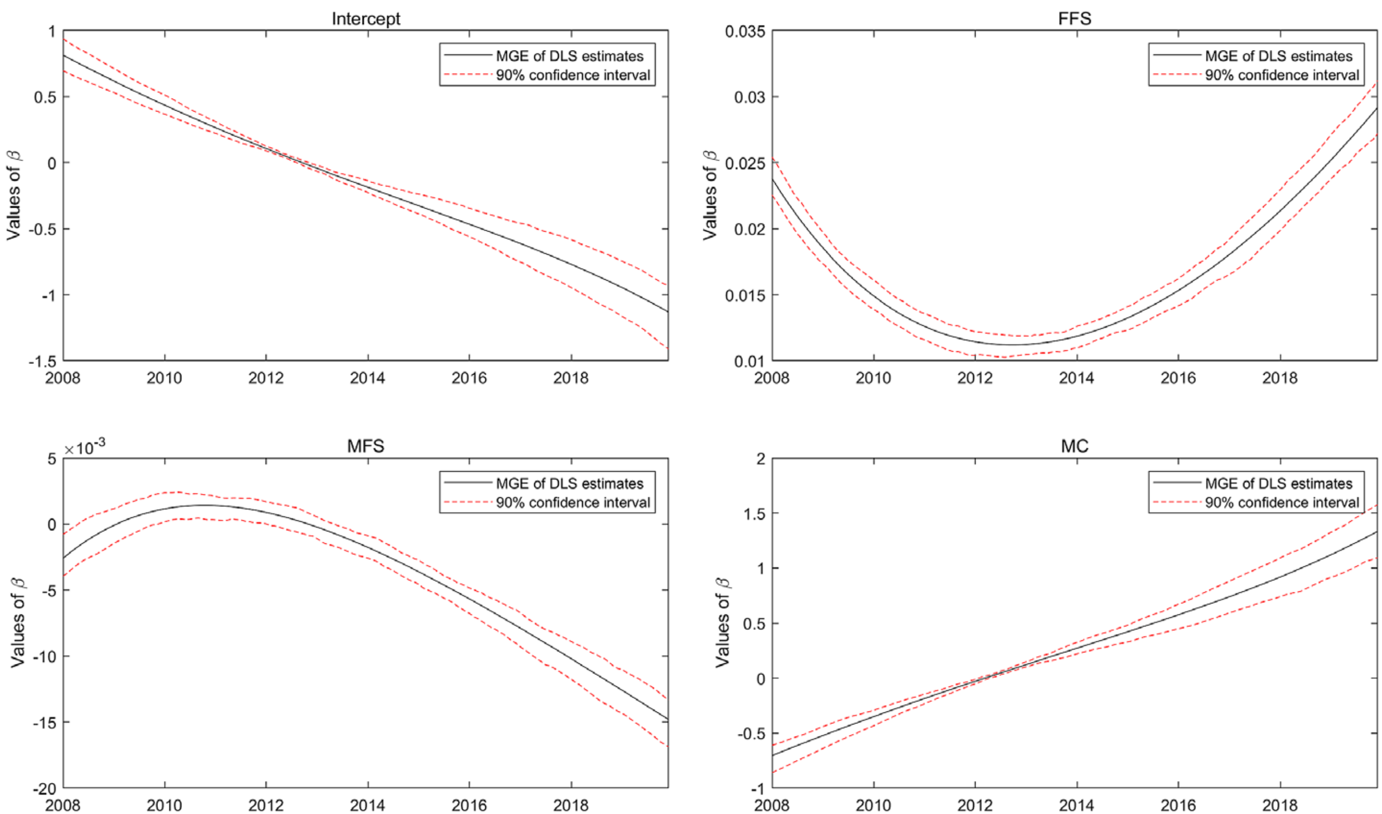


**FIGURE 2.** Estimated MG estimates and 90% confidence intervals.

bias reduction, one may use an undersmoothing bandwidth (e.g., $h \propto T^{-1/3}$) instead. In the initial estimation, we select $\phi_{NT} = (NT)^{-1/10}$ to construct the nuclear-norm penalization estimates. The information criterion method developed in Section 4.2 is then applied to determine the number of factors, for which we specify $r_{max} = 10$ and $d_{NT} = (NT)^{-1/5}$. The corresponding values of IC($r$) for $0 \leq r \leq r_{max}$ are listed in Table 5. We select the number of factors to be 3 according to this information criterion.

After obtaining the DLS estimates for the individual time-varying regression coefficients, we compute the MG estimates using the weight $w_{N,i} = 1/N$ and construct their 90% confidence intervals. As shown in Figure 2, time variations are evident in the associations between U.S. mutual fund returns and fund characteristic variables. The average effect of $\text{FFS}_{it}$ on mutual fund returns is positive and significant, indicating that larger fund companies generally outperform smaller ones in this estimation period. Additionally, we observe from the estimates for $\text{MFS}_{it}$ that fund returns are, on average, negatively related to fund size, except that the effects are insignificant in some periods. Therefore, mutual funds that manage larger assets are typically unable to produce higher returns. This result is in line with Blake et al.'s (2014) findings on mutual funds in the United Kingdom. Our estimation results suggest a time-varying relationship between fund management charges and fund performance, wherein the average effect is negative and significant at the beginning of this estimation period. However, after 2010, we observe the manifestation of an upward trend, which becomes positive thereafter. As presented in Figure 2, significant time-varying patterns exist in all MG estimates. The estimates for $\text{FFS}_{it}$ and $\text{MFS}_{it}$ exhibit quadratic forms, whereas those for the intercept and $\text{MC}_{it}$ are close to linear functions.

Additionally, the constancy test is conducted to further reveal the time-varying relationship between mutual fund characteristics and fund returns. Following the instructions in Section 4.1, we obtain the test statistic's value as $\widetilde{L}_{NT}^{*} = 16.18$ and the corresponding bootstrap $p$-value that is less than 0.01 after 1,000 bootstraps. Therefore, the null hypothesis of time-constant regression coefficients is rejected. We acknowledge that the proposed constancy test requires m.d.s. errors, which can be restrictive in empirical studies. However, allowing for both cross-sectional dependence and temporal correlation involves considerable technical difficulties in the asymptotics and is left as a topic for future research.

## 7. CONCLUSIONS

This paper proposes the DLS iteration method to estimate a panel data model with both heterogeneous time-varying regression coefficients ($\beta_i(\tau)$) and IFEs. Specifically, we iteratively estimate $\beta_i(\tau)$ and factor loadings using the PLS method, and update factor estimators using the OLS method after employing a nuclear-norm penalization approach for initial estimation. This methodology is robust to the correlations between factors and explanatory variables and underlying

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

deterministic trends in the regressors. Under regularity conditions, we establish asymptotic consistency and normality results for the DLS estimators, which can demonstrate the effectiveness of iterations. Additionally, we develop a test procedure to make inferences on the time-constant parameter assumption and propose an information criterion method for factor number selection in practice. The finite-sample performance of the DLS estimation method, constancy test statistic, and information criterion for the factor number selection is assessed through extensive simulations. An empirical study of U.S. mutual funds, which allows for heterogeneity and time variations in parameters, reveals that smaller mutual funds managed by large fund companies can, on average, outperform the others.

## Appendix A. Assumption Justification and Theoretical Development Outline

### A.1. Justifications for Assumptions

**Justification for Assumption 1:** In Assumption 1(i), conditional $\alpha$-mixing conditions are used to restrict the dependence of random errors. Similar to Assumption A.2 from Su and Chen (2013), it can be considered as an extension of the unconditional $\alpha$-mixing condition used by Dong, Gao, and Peng (2015) and Feng et al. (2019). This assumption rules out dynamic panel models, which can be weakened to Assumption A.2 from Su and Chen (2013) to address the problem. The condition regarding the spectral norm of $\mathcal{E}$ is in line with Assumption (ii) in Theorem 1 of Moon and Weidner (2018) and Assumption A.1(vi) from Su and Chen (2013). Assumption 1(iii) is commonly used in the nonparametric estimation literature (e.g. Condition 2.1 from Li and Racine, 2007 and Assumption A3 from Chen, Gao, and Li, 2012). Assumption 1(v) can be justified by the case wherein $N/T \to c_1$, $h = c_2 T^{-1/5}$, and $\phi_{NT} = (NT)^{-1/10}$, where $(c_1, c_2)$ denote fixed constants. Assumption 1(vi) assumes restricted strong convexity, which is a heterogeneous modification of Assumption 1 from Moon and Weidner (2018).

**Justification for Assumption 2:** Assumption 2(i) guarantees the nonsingularity of $\Sigma_f^{\mathcal{D}}$, $\Sigma_{x,i}^{\mathcal{D}}(\tau)$, and $\Omega_{f,i}^{\mathcal{D}}$ and rules out the perfect multicollinearity between $x_{it}$ and $\lambda_i^0, f_t^0$. Assumption 2(ii) is imposed for the positive definiteness of iterative DLS estimators' asymptotic covariance matrices. Moment conditions in Assumption 3(iii) can be justified by conditionally i.i.d. or $\alpha$-mixing processes. Assumption 2(iv) automatically holds for the exponential $\alpha$-mixing process ($\theta \to \infty$) in Assumption 3.4 from Fan, Liao, and Wang (2016).

**Justification for Assumption 3:** Assumption 3(i) assumes that $\Sigma_f^{\mathcal{D}}$, $\Omega_{f,i}^{\mathcal{D}}$, and $\Sigma_{x,i}^{\mathcal{D}}(\tau)$ can be approximated by deterministic matrices $\Sigma_f$, $\Omega_{f,i}$, and $\Sigma_{x,i}(\tau)$, almost surely, which guarantees the asymptotic normalities of estimators. Assumptions 3(ii)–(iv) assumes CLTs, which can be easily justified when $\{v_{it}, \lambda_i^0, f_t^0, \varepsilon_{it}\}$ are i.i.d. across $i$ and $t$.

**Justification for Assumption 4:** Assumption 4 imposes a time-varying version of the random coefficient model in Assumption 4 of Pesaran (2006). Weaker restrictions on $\epsilon_i(\tau)$ could be used instead of the i.i.d. assumption, such as $\alpha$-mixing conditions across $i$. This aspect is left for future research.

https://doi.org/10.1017/S0266466623000324 Published online by Cambridge University Press

**Justification for Assumption 5:** Assumption 5(i) enables us to use the CLT for the $U$-statistic to assess $L_{NT}$'s asymptotic distribution. Assumption 5(ii) further restricts the cross-sectional dependence of random errors, which is necessary for the consistent estimation of $L_{NT}$'s asymptotic variance.

## A.2. Outline of Theoretical Development

This appendix outlines the theoretical development strategy.

First, we study the asymptotic properties of the initial estimators in Theorem 3.1 by demonstrating that minimizing the nonparametric local linear objective function with nuclear-norm regularization ensures consistency under the restricted strong convexity condition. The rates of convergence are derived after we formulate the local linear approximation and regularization estimation bias terms for these estimators.

Building on the initial estimation, Theorem 3.2 establishes the DLS estimators' consistency. By linking the bias terms' probability orders with the number of iterations ($n$), we obtain the preliminary rates of convergence for the DLS estimators and demonstrate the effectiveness of the proposed iteration algorithm.

Furthermore, we study DLS estimators by deriving their leading terms that contribute to the CLTs and establish the asymptotic distributions of the individual and MG estimators in Theorems 3.3 and 3.4, respectively. Thus, we complete the development of the DLS estimators' asymptotic properties.

Thereafter, a residual-based statistic ($\check{L}_{NT}$) is proposed to test the time-constant parameter assumption. By formulating the leading term of $\check{L}_{NT}$ into a $U$-statistic, we employ the CLT for m.d.s. $U$-statistics to establish the asymptotic normality of the test statistic in Theorem 4.1. Theorem 4.2 can be considered as a bootstrap version of Theorem 4.1.

Finally, we propose an information criterion (IC($r$)) that has the same spirit as Bai and Ng's (2002) method for factor number selection and establishes its consistency in Theorem 4.3.

## SUPPLEMENTARY MATERIAL

Fei Liu (2023): Supplement to "Nonparametric Time-Varying Panel Data Models with Heterogeneity," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466623000324

### *References*

Atak, A., O. Linton, & Z. Xiao (2011) A semiparametric panel model for unbalanced data with application to climate change in the United Kingdom. *Journal of Econometrics* 164 (1), 92–115.

Bai, J. (2009) Panel data models with interactive fixed effects. *Econometrica* 77 (4), 1229–1279.

Bai, J. & K. Li (2012) Statistical analysis of factor models of high dimension. *Annals of Statistics* 40 (1), 436–465.

Bai, J. & S. Ng (2002) Determining the number of factors in approximate factor models. *Econometrica* 70 (1), 191–221.

Blake, D., T. Caulfield, C. Ioannidis, & I. Tonks (2014) Improved inference in the evaluation of mutual fund performance using panel bootstrap methods. *Journal of Econometrics* 183 (2), 202–210.

Cai, Z. (2007) Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* 136 (1), 163–188.

Chen, J., J. Gao, & D. Li (2012) Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics* 171 (1), 71–85.

Chernozhukov, V., C. Hansen, Y. Liao, & Y. Zhu (2018). Inference for heterogeneous effects using low - rank estimation of factor slopes. Preprint. https://arxiv.org/pdf/1812.08089.pdf.

Dong, C., J. Gao, & B. Peng (2015) Semiparametric single-index panel data models with cross-sectional dependence. *Journal of Econometrics* 188 (1), 301–312.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Taylor & Francis.

Fan, J., Y. Liao, & W. Wang (2016) Projected principal component analysis in factor models. *Annals of Statistics* 44 (1), 219–254.

Feng, G., B. Peng, L. Su, & T.T. Yang (2019) Semi-parametric single-index panel data models with interactive fixed effects: Theory and practice. *Journal of Econometrics* 212 (2), 607–622.

Ferson, W.E. & R.W. Schadt (1996) Measuring fund strategy and performance in changing economic conditions. *Journal of Finance* 51 (2), 425–461.

Gao, J. & I. Gijbels (2008) Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association* 103 (484), 1584–1594.

Hardle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

Jiang, B., Y. Yang, J. Gao, & C. Hsiao (2021) Recursive estimation in large panel data models: Theory and practice. *Journal of Econometrics* 224 (2), 439–465.

Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92 (1), 101–147.

Li, Q. & J.S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Ma, S., L. Su, & Y. Zhang (2022) Detecting latent communities in network formation models. *Journal of Machine Learning Research* 23 (1), 13971–14031.

Mamaysky, H., M. Spiegel, & H. Zhang (2008) Estimating the dynamics of mutual fund alphas and betas. *Review of Financial Studies* 21 (1), 233–264.

Miao, K., P.C.B. Phillips, & L. Su (2023) High-dimensional VARs with common factors. *Journal of Econometrics* 233 (1), 155–183.

Moon, H.R. & B. Perron (2004) Testing for a unit root in panels with dynamic factors. *Journal of Econometrics* 122 (1), 81–126.

Moon, H.R. & M. Weidner (2018). Nuclear norm regularized estimation of panel regression models. Preprint. https://arxiv.org/pdf/1810.10987.pdf.

Pesaran, M.H. (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74 (4), 967–1012.

Phillips, P.C.B. & Y. Wang (2022) Functional coefficient panel modeling with communal smoothing covariates. *Journal of Econometrics* 227 (2), 371–407.

Speckman, P. (1988) Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B* 50 (3), 413–436.

Su, L. & Q. Chen (2013) Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory* 29 (6), 1079–1135.

Su, L., S. Jin, & Y. Zhang (2015) Specification test for panel data models with interactive fixed effects. *Journal of Econometrics* 186 (1), 222–244.

Su, L. & A. Ullah (2006) Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters* 92 (1), 75–81.

Su, L. & X. Wang (2017) On time-varying factor models: Estimation and testing. *Journal of Econometrics* 198 (1), 84–101.

Wooldridge, J.M. (1992) A test for functional form against nonparametric alternatives. *Econometric Theory* 8 (4), 452–475.